CHAPTER 2

The Cauchy-Kovalevskaya Theorem

We shall start with a discussion of the only "general theorem" which can be extended from the theory of ODE's, the Cauchy-Kovalevskaya Theorem, as it allows to introduce the notion of *principal symbol* and *non-characteristic data* and it is important to see from the start why analyticity is not the proper regularity for studying PDE's most of the time.

1. The Cauchy-Kovalevskaya theorem for ODE's

1.1. Scalar ODE's. As a warm up we will start with the corresponding result for ordinary differential equations.

THEOREM 1.1 (ODE Version of Cauchy–Kovalevskaya, I). *Suppose $a > 0$ and $F : (a,a) \to \mathbb{R}$ is real analytic near 0 and $u(t)$ is the unique solution to the ODE*

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = F(u(t)) \quad \textit{with} \quad u(0) = 0. \tag{1.1}$$

Then u is also real analytic near 0.

We will give four proofs. However it is the last proof that the reader should focus on for understanding the PDE version of Theorem 3.1. Observe that the existence and uniqueness of solutions is granted by Picard-Lindelöf theorem. However existence and uniqueness arguments could be devised using the arguments below as well. Observe also that it is enough to show the regularity in a neighbourhood of 0, even small, as then the argument can be performed again around any point where u is defined and F is analytic.

PROOF 1 OF THEOREM 1.1. We follow here the same strategy as for solving an ODE by "separation of variables". If $F(0) = 0$ then the solution is $u = 0$ which is clearly analytic and we are done. Assume $F(0) \neq 0$, then let us define the new function

$$G(y) = \int_0^y \frac{1}{F(x)}\,\mathrm{d}x, \quad y \in (-a', a') \subset (-a, a)$$

which is again real analytic in a neighbourhood of 0. Then we have by the chain rule in the (possibly smaller) neighbourhood where u is defined and

G is analytic:

$$\frac{\mathrm{d}}{\mathrm{d}t} G(u(t)) = \frac{\dot{u}(t)}{F(u(t))} = 1$$

which implies, together with $G(u(0)) = G(0) = 0$, that $G(u(t)) = t$. Then $u(t) = G^{-1}(t)$ is analytic near 0 as $G'(0) = 1/F(0) \neq 0$. □

PROOF 2 OF THEOREM 1.1. Let us consider, for $z \in CC$, the solution $u_z(t)$ to

$$\dot{u}_z(t) = zF(u_z(t)). \tag{1.2}$$

Observe that if u is analytic, it extends to neighbourhood in the complex plane, and satisfies there the equation (1.2). One can show by calculations that for any solution to (1.2) one has

$$\frac{\partial}{\partial t} \frac{\partial u_z(t)}{\partial \bar{z}} = zf'(u_z(t)) \frac{\partial u_z(t)}{\partial \bar{z}}$$

which implies, together with the initial condition $u_z(0) = 0$, that $\partial u_z(t)/\partial \bar{z} = 0$ in the t-neighbourhood of 0, which are the Cauchy-Riemann equations. Let us recall the notation

$$\frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right), \quad \frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right).$$

One then construct by Picard-Lindelöf a solution to (1.2) on a neighbourhood $|t| \le \varepsilon$ and $|z| \le 2$ (the Lipshitz influences the size of the neighbourhood), and since $z \mapsto u_z(t)$ satisfies the Cauchy-Riemann equations in this neighbourhood we deduce

$$u_1(t) = \sum_{n=0}^{\infty} \frac{1^n}{n!} \left(\frac{\partial^n u_z(t)}{\partial z^n} \right) \Big|_{z=0}.$$

Now we use that for $z \in \mathbb{R}$ **real** we always have $u(t) = u(zt)$, which implies that

$$\left(\frac{\partial^n u_z(t)}{\partial z^n} \right) \Big|_{z=0} = \left(\frac{\partial^n u(zt)}{\partial z^n} \right) \Big|_{z=0} = t^n u^{(n)}(0)$$

which yields

$$u(t) = u_1(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} u^{(n)}(0)$$

and shows the real analyticity of u. □

PROOF 3 OF THEOREM 1.1. This proof is left as an exercise: go back to the proof of Picard-Lindelöf theorem by fixed-point argument, and replace the real variable by a complex variable and the real integral by a complex path integral:

$$u_{n+1}(z) = \int_0^z F(u_n(z'))\,\mathrm{d}z' = \int_0^1 F(u_n(zt))z\,\mathrm{d}t$$

and show that the contraction property and the fixed point can be performed in the space of holomorphic functions. □

PROOF 4 OF THEOREM 1.1. This is the most important proof, as it is the historic proof of A. Cauchy (improved by S. Kovalevskaya) but also because it is beautiful and this is the proof we shall use in a PDE context. This is called the "method of majorants". Let us do first an *a priori examination of the problem*, assuming the analyticity (actually here it could be justified by using Picard-Lindelöf to contruct solutions, and then check by boostrap that this solution is C^∞). Then we compute the derivatives

$$\begin{cases} u^{(1)}(t) = F^{(0)}(u(t)), \\ u^{(2)}(t) = F^{(1)}(u(t))u^{(1)}(t) = F^{(1)}(u(t))f^{(0)}(u(t)), \\ u^{(3)}(t) = F^{(2)}(u(t))f(u(t))^2 + F^{(1)}(u(t))^2 F(u(t))\dots \end{cases}$$

REMARK 1.2. *The calculation of these polynomials is connected to a formula devised in the 19th century by L. Arbogast in France and Faà di Bruno in Italy. It is now known as Faà di Bruno's formula and it is good to keep it in one's analytic toolbox:*

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n}F(u(t)) = \sum_{m_1+2m_2+\cdots+nm_n=n} \frac{n!}{m_1!1!^{m_1}m_2!2!^{m_2}\dots m_n n!^{m_n}} F^{(m_1+\cdots+m_n)}(u(t)) \prod_{j=1}^n \left(u^{(j)}(t)\right)^{m_j}.$$

Now the key observation is that there are universal (in the sense of being independent of the function F) polynomials p_n with **non-negative** integer coefficients, so that

$$u^{(n)}(t) = p_n\left(F^{(0)}(u(t)), \dots, F^{(n-1)}(u(t))\right).$$

We deduce by monotonicity

$$|u^{(n)}(0)| \le p_n\left(|F^{(0)}(0)|, \dots, |F^{(n-1)}(0)|\right) \le p_n\left(G^{(0)}(0), \dots, G^{(n-1)}(0)\right)$$

for any function G with **non-negative** derivatives at zero and such that $G^{(n)}(0) \ge |f^{(n)}(0)|$ for all $n \ge 0$. Such a function is called a *majorant* function of F.

But for such a function G the RHS in the previous equation is exactly

$$p_n\left(G^{(0)}(0), \ldots, G^{(n-1)}(0)\right) = v^{(n)}(0) = |v^{(n)}(0)|$$

where v solves the auxiliary equation

$$\frac{\mathrm{d}}{\mathrm{d}t} v(t) = G(v(t)).$$

Hence if v is analytic near zero, the series

$$\mathcal{S}_v(t) := \sum_{n\ge 0} v^{(n)}(0) \frac{t^n}{n!}$$

has a positive radius of convergence and by comparison so does the series

$$\mathcal{S}_u(t) := \sum_{n\ge 0} |u^{(n)}(0)| \frac{t^n}{n!},$$

which shows the analyticity near zero and concludes the proof.

We finally need to construct the majorant function G. We use the analyticity of F to deduce readily (comparing series...)

$$\forall\, n \ge 0, \quad |F^{(n)}(0)| \le C \frac{n!}{r^n}$$

uniformly in $n \ge 0$, for some some constant $C > 0$ and some $r > 0$ smaller than the radius of convergence of the series, and we then consider

$$G(z) := C \sum_{n=0}^{\infty} \left(\frac{z}{r}\right)^n = C \frac{1}{1 - z/r} = \frac{Cr}{r - z}$$

which is analytic on the ball centred at zero with radius $r > 0$. Since $G^{(n)}(0) = Cn!/r^n$ we have clearly the majoration $G^{(n)}(0) \ge |f^{(n)}(0)|$ for all $n \ge 0$.

To conclude the proof we need finally to compute the solution v to the auxiliary equation

$$\frac{\mathrm{d}}{\mathrm{d}t} v(t) = G(v(t)) = \frac{Cr}{r - v(t)}, \quad v(0) = 0,$$

which can be solved by usual real differential calculus, using separation of variables:

$$(r-v)\,\mathrm{d}v = Cr\,\mathrm{d}t \quad \Longrightarrow \quad -\mathrm{d}(r-v)^2 = Cr\,\mathrm{d}t \quad \Longrightarrow \quad v(t) = r \pm r\sqrt{1 - \frac{2Ct}{r}}$$

and using the initial condition $v(0) = 0$ one finally finds

$$v(t) = r - r\sqrt{1 - \frac{2Ct}{r}}$$

which is analytic for $|t| \leq r/(2C)$. This hence shows that the radius of convergence of $\mathcal{S}_v(t)$ is positive, which concludes the proof. □

Observe the profound idea in this last proof: instead of coping with the combinatorial explosion in the calculation of the derivatives (due to the nonlinearity), one uses a monotonicity property encoded in the abstract structure, reduce the control to be established to a comparison with a simpler function, and then go back to the equation without computing the derivatives anymore.

1.2. Systems of ODE's. We now consider the extension of this theorem to *systems* of differential equations.

THEOREM 1.3 (ODE Version of Cauchy–Kovalevskaya, II). *Suppose* $a > 0$ *and* $\mathbf{F} : (a, a)^m \to \mathbb{R}^m$, $m \in \mathbb{N}$, *is real analytic near* $0 \in (-a, a)^m$ *and* $\mathbf{u}(t)$ *is the unique solution to the ODE*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}(t) = \mathbf{F}(\mathbf{u}(t)) \quad \textit{with} \quad \mathbf{u}(0) = 0. \tag{1.3}$$

Then $\mathbf{u}$ *is also real analytic near* 0.

PROOFS OF THEOREM 1.3. All but the first proof of Theorem 1.1 can be adapted to cover this case of systems.

EXERCISE 10. *Extend the proofs 2 and 3 to this case.*

Let us give some more comments-exercises on the extension of the last proof, the method of majorants.

EXERCISE 11. *Suppose* $\mathbf{F} : (-a, a)^m \to \mathbb{R}^m$ *is real analytic near* $0 \in (-a, a)^m$*, prove that a majorant function is provided by*

$$\mathbf{G}(z_1, \dots, z_m) := (G_1, \dots, G_m), \quad G_1 = \dots = G_m = \frac{Cr}{r - z_1 - \dots - z_m}$$

for well-chosen values of the constants $r, C > 0$.

With this auxiliary result at hand, check that one can reduce the proof to proving the local analyticity of the solution to the system of ODE:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}(t) = \mathbf{G}(\mathbf{v}(t)), \quad \mathbf{v}(t) = (v_1(t), \dots, v_m(t)), \quad \mathbf{v}(0) = 0.$$

EXERCISE 12. *Prove that by symmetry one has* $v_j(t) = v_1(t) =: w(t)$ *for all* $1 \le j \le m$*, and that* $w(t)$ *solves the* scalar *ODE*

$$\frac{\mathrm{d}}{\mathrm{d}t} w(t) = \frac{Cr}{r - mw(t)}, \quad w(0) = 0$$

so that $w(t) = (r/m)(1 - \sqrt{1 - 2Cdt/r})$.

With the two last results it is easy to conclude the proof. □

2. The analytic Cauchy problem in PDE's

We consider a k-th order scalar *quasilinear* PDE[1]

(2.1)
$$\sum_{|\alpha|=k} a_\alpha(\nabla^{k-1}u, \dots, u, x)\partial_x^\alpha u + a_0(\nabla^{k-1}u, \dots, u, x) = 0, \quad x \in \mathcal{U} \subset \mathbb{R}^\ell$$

where

$$\nabla^l u := \Big(\partial_{x_{i_1}} \dots \partial_{x_{i_l}} u\Big)_{1 \le i_1, \dots, i_l \le \ell}, \quad l \in \mathbb{N},$$

is the l-th iterated gradient, and

$$\partial_x^\alpha := \partial_{x_1}^{\alpha_1} \dots \partial_{x_\ell}^{\alpha_\ell}$$

for a multi-index $\alpha = (\alpha_1, \dots, \alpha_\ell) \in \mathbb{N}^\ell$, and $\mathcal{U}$ is some open region in $\mathbb{R}^\ell$ ($\ell \ge 2$ is the number of variables), and $u : \mathcal{U} \to \mathbb{R}$.

REMARK 2.1. *The word "quasilinear" relates to the fact that the coefficient of the highest-order derivatives only depend on derivatives with strictly lower order. The equation would be* semilinear *if* $a_\alpha = a_\alpha(x)$ *does not depend on* u *and the nonlinearity is only in* a_0*. The equation is* linear *when of course both* a_α *and* a_0 *do not depend on* u*, and it is a* constant coefficient linear equation *when* a_α *and* a_0 *do not depend on* x *either.*

We consider a smooth $(\ell-1)$-dimensional hypersurface Γ in $\mathcal{U}$, equiped with a smooth unit normal vector $\mathbf{n}(x) = (n_1(x), \dots, n_\ell(x))$ for $x \in \Gamma$. We then define the *j-th normal derivative* of u at $x \in \Gamma$

$$\frac{\partial^j u}{\partial \mathbf{n}^j} := \sum_{|\alpha|=j} n^\alpha \partial_x^\alpha u = \sum_{\alpha_1 + \dots \alpha_\ell = j} \frac{\partial^j u}{\partial x_1^{\alpha_1} \dots \partial x_\ell^{\alpha_\ell}} n_1^{\alpha_1} \dots n_\ell^{\alpha_\ell}.$$

[1]Actually when $k = 1$ (first order) another simpler proof than the one we shall do here can be performed by the so-called *characteristics method*. This method can be understood as the natural extension of the ODE arguments in proofs 1-2-3 above using trajectories for the PDE. However this method fails for systems, and therefore is unable to treat k-th order PDE's as we shall see, which justifies the need for a more general proof.

Now let $g_0, \ldots, g_{k-1} : \Gamma \to \mathbb{R}$ be k given functions on Γ. The *Cauchy problem* is then to find a function u solving (2.1), subject to the boundary conditions

$$u = g_0, \quad \frac{\partial u}{\partial \mathbf{n}} = g_1, \quad \ldots \quad , \quad \frac{\partial^{k-1} u}{\partial \mathbf{n}^{k-1}} = g_{k-1}, \quad x \in \Gamma. \tag{2.2}$$

We say that the equation (2.2) prescribes the *Cauchy data* g_0, ..., g_{k-1} on Γ.

If one wants to compute an entire series for the solution, certainly all the derivatives have to be determined from equations (2.1)-(2.2). In particular *all* partial derivatives of u on Γ should be computed from the boundary data (2.2). The basic question is now: what kind of conditions do we need on Γ in order to so?

2.1. The case of a flat boundary. In order to gain intuition into the problem, we first examine the case where $\mathcal{U} = \mathbb{R}^\ell$ and $\Gamma = \{x_\ell = 0\}$ is a vector hyperplan. We hence have $\mathbf{n} = \mathbf{e}_\ell$ (the ℓ-th unit vector of the canonical basis) and the boundary prescriptions (2.2) read

$$u = g_0, \quad \frac{\partial u}{\partial x_\ell} = g_1, \quad \ldots \quad , \quad \frac{\partial^{k-1} u}{\partial x_\ell^{k-1}} = g_{k-1}, \quad x \in \Gamma.$$

Which further partial derivatives can we compute on the hyperplan Γ? First since $u = g_0$ on Γ by differentiating tangentially we get that

$$\frac{\partial u}{\partial x_i} = \frac{\partial g_0}{\partial x_i}, \quad 1 \le i \le \ell - 1$$

is prescribed by the boundary data. Since we also know from (2.2) that

$$\frac{\partial u}{\partial x_\ell} = g_1$$

we can determine the full gradient on Γ. Similarly we can calculate inductively

$$\frac{\partial^\alpha}{\partial \mathbf{x}^\alpha} \frac{\partial^j u}{\partial x_\ell^j} = \frac{\partial^\alpha}{\partial \mathbf{x}^\alpha} g_j, \quad \alpha = (\alpha_1, \ldots, \alpha_{\ell-1}, 0), \ |\alpha_i| \le k - 1, \ 0 \le j \le k - 1.$$

Remark that actually the α_i in the previous equation could be taken in $\mathbb{N}$. The difficulty now, in order to compute the k-th derivative, is to compute the k-th order normal derivative

$$\frac{\partial^k u}{\partial x_\ell^k}$$

We now shall use the PDE in order to overcome this obstacle. Observe that if the coefficient a_α with $\alpha = (0,\dots,0,k)$ is non-zero on Γ:

$$\begin{aligned} A(x) &:= a_{(0,\dots,0,k)}(\nabla^{k-1}u,\dots,u,x) \\ &= \mathrm{Function}(g_{k-1}(x), g_{k-2}(x),\dots,g_0(x),x) \neq 0, \quad x \in \Gamma, \end{aligned}$$

(observe that it only depends on the boundary data) then we can compute for $x \in \Gamma$

$$\frac{\partial^k u}{\partial x_\ell^k} = -\frac{1}{A(x)}\left[\sum_{|\alpha|=k,\ \alpha_\ell \le k-1} a_\alpha(\nabla^{k-1}u,\dots,u,x)\partial_x^\alpha u + a_0(D^{k-1}u,\dots,u,x)\right]$$

where the coefficients in the RHS again only depend on the boundary data by the previous calculations, and consequently we can therefore compute $\nabla^k u$ on Γ.

We say therefore that the surface $\Gamma = \{x_\ell = 0\}$ is *non-characteristics* for the PDE (2.1) if the function $A(x) = a_{(0,\dots,0,k)}$ never cancels on Γ.

Now the question is: can we calculate still higher derivatives on Γ, assuming of course this non-degeneracy condition? The answer is yes, here is a concise inductive way of iterating the argument:

Let us denote

$$g_k(x) := \frac{\partial^k u}{\partial x_\ell^k} = -\frac{1}{A(x)}\left[\sum_{|\alpha|=k,\ \alpha_\ell \le k-1} a_\alpha \partial_x^\alpha u + a_0\right], \quad x \in \Gamma,$$

as computed before. We now differentiate the equation along x_ℓ (we already know how to compute all the derivatives along the other coordinates, provided we have less than k derivatives along x_ℓ), which results into a new equation

$$\sum_{|\alpha|=k} a_\alpha(\nabla^{k-1}u,\dots,u,x)\partial_x^\alpha \partial_{x_\ell} u + \tilde{a}_0(\nabla^k u,\dots,u,x) = 0, \quad x \in U \subset \mathbb{R}^\ell,$$

which results following the same argument into

$$\frac{\partial^{k+1} u}{\partial x_\ell^{k+1}} = -\frac{1}{A(x)}\left[\sum_{|\alpha|=k,\ \alpha_\ell \le k-1} a_\alpha(\nabla^{k-1}u,\dots,u,x)\partial_x^\alpha \partial_{x_\ell} u + \tilde{a}_0(\nabla^k u,\dots,u,x)\right]$$

(observe that the RHS only involves derivatives in x_ℓ of order less than k), which allows to calculate the $k+1$-derivative in x_ℓ from the boundary data. One can then continue inductively and calculate all derivatives.

2.2. General surfaces. We shall now generalize the results and definitions above to the general case, when Γ is a smooth hypersurface with normal vector field $\mathbf{n}$.

DEFINITION 2.2. *We say that the surface* Γ *is* non-characteristic *for the PDE* (2.1) *if*

$$A(x) := \sum_{|\alpha|=k} a_\alpha \mathbf{n}^\alpha \neq 0, \quad x \in \Gamma$$

(where the RHS only depends on the boundary data).

Let us prove the theorem corresponding the calculation of the partial derivatives

THEOREM 2.3 (Cauchy data and non-characteristic surfaces). *Assume that* Γ *is* C^{∞}[2] *and non-characteristic for the PDE* (2.1)*. Then if* u *is a* C^∞ *solution to* (2.1) *with the boundary data* (2.2)*, we can uniquely compute all the partial derivatives of* u *on* Γ *in terms of* Γ*, the functions* $g_0, \ldots, g_{k-1}$*, and the coefficients* a_α*,* a_0*.*

PROOF OF THEOREM 2.3. We consider a base point $x \in \Gamma$, and using the smoothness of the Γ we find C^∞ maps Φ, Ψ defined on open sets of $\mathbb{R}^\ell$ to $\mathbb{R}^\ell$ so that

$$\Phi(\Gamma \cap B(x,r)) = \Theta \subset \{y_\ell = 0\}, \quad \Phi(x) = y, \quad \Psi = \Phi^{-1}$$

for some $r > 0$, where Θ is the new Cauchy surface in the new coordinates, and with the property

$$\frac{\partial \Psi}{\partial y_\ell}(y) = \lambda(y)\mathbf{n}(y) \quad \text{and} \quad \mathbf{n} = \mathbf{n}(y_1, \ldots, y_{\ell-1}) \text{ does not depend on } y_\ell,$$

for some $\lambda(y) \neq 0$ on Θ: this means for instance that

$$\Psi(\bar{y}, y_\ell) = \bar{\Psi}(\bar{y}) + y_\ell \mathbf{n}(\bar{y}) \quad \text{with} \quad \bar{y} = (y_1, \ldots, y_{\ell-1}) \quad \text{and} \quad \bar{\Psi}(\bar{y}) \in \Gamma.$$

Then we define

$$v(y) := u(\Psi(y))$$

and it is a straightforward calculation to show that v satisfies a new equation of the form

$$\sum_{|\alpha|=k} b_\alpha(\nabla_y^{k-1} v, \ldots, v, y)\partial_y^\alpha v + b_0(\nabla_y^{k-1} u, \ldots, u, y) = 0, \quad y \in \mathcal{V} \subset \mathbb{R}^\ell \tag{2.3}$$

[2]This means defined by a C^∞ function.

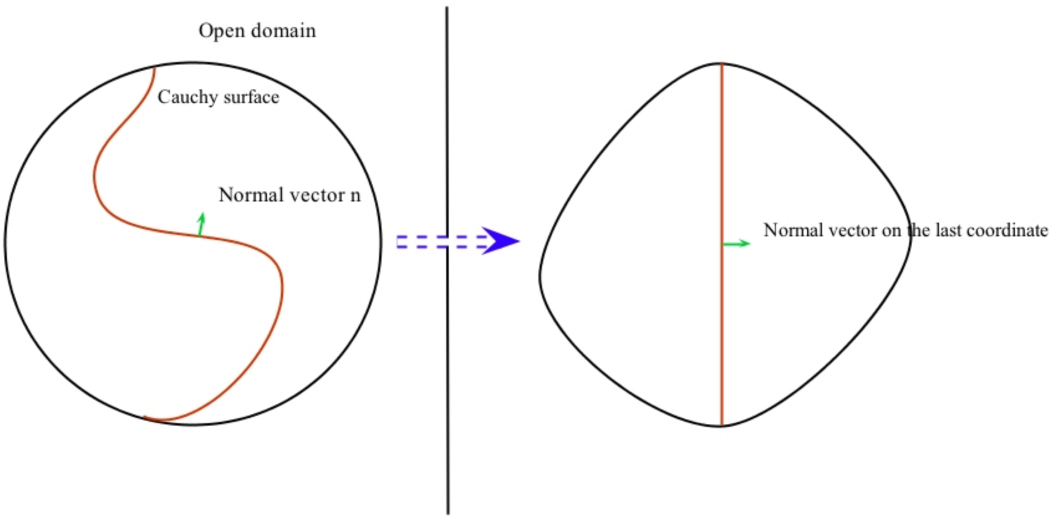

FIGURE 1. Local rectification of the flow.

for some open set $\mathcal{V} \subset \mathbb{R}^\ell$. The new boundary data are

$$\begin{cases} v(y) = g_0(\Psi(y)) =: h_0(y), \\ \dfrac{\partial v}{\partial y_\ell}(y) = (\nabla_x u)(\Psi(y))\dfrac{\partial \Psi}{\partial y_\ell}(y) \\ \qquad\qquad = \lambda_\Psi(y)\dfrac{\partial u}{\partial \mathbf{n}}(\Psi(y)) = \lambda_\Psi(y) g_1(\Psi(y)) =: h_1(y), \\ \dfrac{\partial^2 v}{\partial y_\ell^2}(y) = (\nabla_x^2 u)(\Psi(y)) : \left(\dfrac{\partial \Psi}{\partial y_\ell}(y)\right)^{\otimes 2} + \dfrac{\partial \lambda_\Psi}{\partial y_\ell} g_1(\Psi(y)) \\ \qquad\qquad = \lambda_\Psi(y)^2 \dfrac{\partial^2 u}{\partial \mathbf{n}^2}(\Psi(y)) + \dfrac{\partial \lambda_\Psi}{\partial y_\ell} g_1(\Psi(y)) \\ \qquad\qquad = \lambda_\Psi(y)^2 g_2(\Psi(y)) + \dfrac{\partial \lambda_\Psi}{\partial y_\ell} g_1(\Psi(y)) =: h_2(y) \\ \dfrac{\partial^3 v}{\partial y_\ell^3}(y) = \cdots =: h_3(y) \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial^{k-1} v}{\partial y_\ell^{k-1}}(y) = \cdots =: h_{k-1}(y) \end{cases}$$

and one checks by induction that they can be computed only in terms of the boundary conditions.

Then we have again the non-characteristic property in the new coordinates on the new Cauchy surface

$$b_{(0,\dots,0,k)}(\nabla_y^{k-1} v, \dots, v, y) \neq 0, \quad y \in \Theta = \mathcal{V} \cap \{y_\ell = 0\}.$$

Indeed we calculate for $\alpha = (0, \dots, 0, k)$ that

$$\frac{\partial^\alpha u}{\partial x^\alpha}(x) = \frac{\partial^k v}{\partial y_\ell^k}(y)\,(\nabla \Phi(x))^\alpha + \quad \text{lower-order terms}$$

where the lower-order terms only involve partial derivatives with order less than $k-1$ in y_ℓ. We hence deduce that

$$b_{(0,\dots,0,k)}(\nabla_y^{k-1} v, \dots, v, y) = \sum_{|\alpha|=k} a_\alpha(\nabla_x^{k-1} u, \dots, u, x)\mathbf{n}^\alpha = A(\Psi(y)) \neq 0.$$

Exercise 13. *Check the previous calculation.*

Then using the previous case of a flat boundary it concludes the proof. □

3. The Cauchy-Kovalevskaya Theorem for PDE's

We shall now prove the following result:

Theorem 3.1 (Cauchy-Kovalevakaya Theorem for PDE's). *Under analyticity assumptions on all coefficients, and the non-characteristic condition, there is a unique local analytic solution u to the equations* (2.1)-(2.2).

This result was first proved by A. Cauchy in 1842 on first order quasilinear evolution equations, and formulated in its most general form by S. Kovalevskaya in 1874. At about the same time, G. Darboux also reached similar results, although with less generality than Kovalevskaya's work. Both Kovalevskaya's and Darboux's papers were published in 1875, and the proof was later simplified by E. Goursat in his influential calculus texts around 1900. Nowadays these results are collectively known as the *Cauchy-Kovalevskaya Theorem.*

3.1. Reduction to a first-order system with flat boundary. We now consider an *analytic*[3] Cauchy surface Γ and the PDE (2.1) with boundary data (2.2), where all the coefficients a_α, a_0, g_0, …, g_{k-1} are analytic in all their variable.[4]

- First, upon flattening out the boundary by an analytic mapping, we can reduce to the situation that $\Gamma = \{x_\ell = 0\}$.

- Second, upon dividing properly by $a_{(0,\dots,0,k)}$ locally around Γ, we can assume that $a_{(0,\dots,0,k)} = 1$ by changing the coefficients to new (still analytic) coefficients.

- Third, by substracting off appropriate analytic functions, we may assume that the Cauchy are identically zero.

- Fourth, we transform the equation to a first-order *system* by defining

$$\mathbf{u} := \left(u, \frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_\ell}, \frac{\partial^2 u}{\partial x_i \partial x_j}, \dots \right)$$

where the vector includes all partial derivatives with total order less than $k-1$. Let m denotes the number of components of this vector. It results into the following system with boundary conditions

$$\left\{ \begin{array}{l} \dfrac{\partial \mathbf{u}}{\partial x_\ell} = \displaystyle\sum_{j=1}^{\ell-1} \mathbf{b}_j(\mathbf{u}, x') \frac{\partial \mathbf{u}}{\partial x_j} + \mathbf{b}_0(\mathbf{u}, x'), \quad x \in \mathcal{U} \\ \\ \mathbf{u} = 0 \quad \text{on} \quad \Gamma, \end{array} \right. \tag{3.1}$$

with matrix-valued functions $\mathbf{b}_j : \mathbb{R}^m \times \mathbb{R}^{\ell-1} \mapsto \mathcal{M}_{m\times m}$ and vector-valued function $\mathbf{b}_0 : \mathbb{R}^m \times \mathbb{R}^{\ell-1} \mapsto \mathbb{R}^m$ which are locally analytic around $(0,0)$, and where $x' = (x_1, \dots, x_{\ell-1})$. Observe that we have assume that the coefficients $\mathbf{b}_j$, $j = 0, \dots, \ell-1$, do not depend on x_ℓ. This can obtained by adding a further component $\mathbf{u}_{m+1} = x_\ell$ if necessary.

Remark 3.2. *Observe that the reduction in this subsection uses crucially the non-characteristic condition. It means at a physical level that we have been able to use one of the variables as a time variable in order to reframe the problem as an evolution problem. However finding a non-characteristic Cauchy surface to start with can be difficult, this is for instance one of the issues in solving the Einstein equations in general relativity, as in the Choquet-Bruhat Theorem.*

[3]In the sense of being implicitly defined by an analytic function, or equivalently locally rectifiable with analytic maps.

[4]Let us recall that for a real function of several variables, this means to be locally equal to the Taylor series in all the variables.

3.2. The proof in the reduced case. We now consider a base point on Γ, say 0 w.l.o.g., and, as in the ODE case, we calculate the partial derivatives at this point by repeatedly differentiating the equation.

We have first obviously $\mathbf{u}(0) = 0$.

Second by differentiating the boundary data in x' we get

$$\partial_x^\alpha \mathbf{u}(0) = 0, \quad \text{for any} \quad \alpha \quad \text{with} \quad \alpha_\ell = 0.$$

Then for α with $\alpha_\ell = 1$ we calculate using the PDE (3.1) (denoting $\alpha' = (\alpha_1, \dots, \alpha_{\ell-1}, 0)$):

$$\partial_x^\alpha \mathbf{u} = \sum_{j=1}^{\ell-1} \partial_x^{\alpha'} \left(\mathbf{b}_j(\mathbf{u}, x') \frac{\partial \mathbf{u}}{\partial x_j} \right) + \partial_x^{\alpha'} \mathbf{b}_0(\mathbf{u}, x')$$

which yields at $x = 0$ (using the previous step):

$$\partial_x^\alpha \mathbf{u}(0) = 0 + \left(\partial_x^{\alpha'} \mathbf{b}_0(\mathbf{u}, x') \right)_{\big| x=0} = \left(D_2^{\alpha'} \mathbf{b}_0 \right)(0,0)$$

where D_2 means the partial derivatives according the second argument of $\mathbf{b}_0$.

Then for α with $\alpha_\ell = 2$ we calculate again (denoting $\alpha' = (\alpha_1, \dots, \alpha_{\ell-1}, 1)$):

$$\partial_x^\alpha \mathbf{u} = \sum_{j=1}^{\ell-1} \partial_x^{\alpha'} \left(\mathbf{b}_j(\mathbf{u}, x') \frac{\partial \mathbf{u}}{\partial x_j} \right) + \partial_x^{\alpha'} \mathbf{b}_0(\mathbf{u}, x')$$

which yields at $x = 0$:

$$\partial_x^\alpha \mathbf{u}(0) = \cdots = \text{polynomial}(\partial \mathbf{b}_j(0,0), \partial \mathbf{u}(0))$$

where is in the RHS it only involves derivatives of $\mathbf{u}$ with $\alpha_\ell \leq 1$, which then can be expressed in terms of derivatives of $\mathbf{b}_j$ again.

We can continue the calculation inductively, and prove by induction that there are universal (independent of $\mathbf{u}$) polynomials with integer non-negative coefficients so that

$$\frac{\partial^\alpha \mathbf{u}_i}{\partial x^\alpha} = p_{\alpha,i}(\text{derivatives of } \mathbf{b}, \mathbf{c} \dots).$$

We perform then the same argument as for ODE's with the majorant function

$$\mathbf{b}_j^* = \frac{Cr}{r - (x_1 + \dots x_{\ell-1}) - (z_1 + \cdots + z_m)} \mathbf{M}_1, \quad j = 1, \dots, \ell - 1,$$

where $\mathbf{M}_1$ is the $m \times m$-matrix with 1 in all entries, and

$$\mathbf{b}_0^* = \frac{Cr}{r - (x_1 + \dots x_{\ell-1}) - (z_1 + \cdots + z_m)} \mathbf{U}_1$$

where $\mathbf{U}_1$ is the m-vector with 1 in all entries, resulting in the solution

$$\mathbf{v} = \frac{1}{m\ell}\left(r - (x_1 + \cdots + x_{\ell-1})\right) - \left[(r - (x_1 + \cdots + x_{\ell-1}))^2 - 2m\ell C r x_\ell\right]^{1/2} \mathbf{U}_1$$

which yields the analyticity in all variables.

To give a proof of the last sentence we shall decompose several steps, given in exercises.

Exercise 14. *Using the exercise 11 on all entries of* $\mathbf{b}_j$, $j = 0, \ldots, \ell-1$ *(which depend on* $m + \ell - 1$ *variables), find* $C, r > 0$ *so that*

$$g(z_1, \ldots, z_m, x_1, \ldots, x_{\ell-1}) = \frac{Cr}{r - (x_1 + \ldots x_{\ell-1}) - (z_1 + \cdots + z_m)}$$

is a majorant of all these entries.

Exercise 15. *Defining* $\mathbf{b}_j^* = g\mathbf{M}_1$, $j = 1, \ldots, \ell - 1$, *and* $\mathbf{b}_0^* = g\mathbf{U}_1$, *check that the solution* $\mathbf{v} = (v_1, \ldots, v_m)$ *to*

$$\begin{cases} \dfrac{\partial \mathbf{v}}{\partial x_\ell} = \displaystyle\sum_{j=1}^{\ell-1} \mathbf{b}_j^*(\mathbf{v}, x') \frac{\partial \mathbf{u}}{\partial x_j} + \mathbf{b}_0^*(\mathbf{v}, x') \\ \mathbf{v} = 0 \quad \text{on} \quad \Gamma, \end{cases}$$

can be searched in the form $v_l = v_1 =: w$, $l = 1, \ldots, m$, *and*

$$w = w(x_1 + x_2 + \cdots + x_{\ell-1}, x_\ell) = w(y, x_\ell), \quad y := x_1 + \cdots + x_{\ell-1}.$$

Then it reduces the problem to solving the following scalar simple transport equation (relabeling $x_\ell = t$ for conveniency)

$$\partial_t w = \frac{Cr}{r - y - \gamma_1 w}\left(\gamma_2 \partial_y w + 1\right), \quad w(0, y) = 0, \quad t, y \in \mathbb{R}. \tag{3.2}$$

Exercise 16. *Show the* w *defining the solution* $\mathbf{v}$ *to the majorant problem above satisfies the equation* (3.2) *with* $\gamma_2 = (\ell - 1)m$ *and* $\gamma_1 = m$.

Finally we can solve the equation (3.2) by the so-called *characteristic method* (which we shall study in much more details in the chapter on hyperbolic equations). Let us sketch the method in this case: if we can find $y(t)$ and $z(t)$ solving

$$\begin{cases} y'(t) = \dfrac{-Cr\gamma_2}{r - y - \gamma_1 z}, \quad y(0) = y_0, \\ z'(t) = \dfrac{Cr}{r - y - \gamma_1 z}, \quad z(0) = z_0, \end{cases}$$

then if we set $z_0 = 0$ and now define $w(t, y)$ by the implicit formula $w(t, y(t)) = z(t)$, it solves by the chain-rule

$$z'(t) = (\partial_t w)(t, y(t)) + y'(t)(\partial_y w)(t, y(t)) = \frac{Cr}{r - y(t) - \gamma_1 z(t)}$$

which writes

$$(\partial_t w)(t, y(t)) - \frac{Cr\gamma_2}{r - y - \gamma_1 w(t, y(t))}(\partial_y w)(t, y(t)) = \frac{Cr}{r - y(t) - \gamma_1 w(t, y(t))}$$

with the initial data $w(0, y_0) = z(0) = 0$. This is exactly the desired equation at the point $(t, y(t))$. Hence as long as the map $y_0 \mapsto y(t)$ is invertible and smooth, we have a solution to the original PDE problem.[5]

Therefore let us solve locally in time the ODE's for $y(t)$ and $z(t)$. Since obviously $y'(t) + \gamma_2 z'(t) = 0$, we deduce the key a priori relation

$$\forall\, t \geq 0, \quad y(t) + \gamma_2 z(t) = y_0.$$

Whe thus replace $y(t)$ in the ODE for $z(t)$:

$$z'(t) = \frac{Cr}{r - y_0 + (\gamma_2 - \gamma_1)z}, \quad z(0) = 0.$$

Here observe that $\gamma_2 \geq \gamma_1$ (remember that $\ell \geq 2$). If $\gamma_1 = \gamma_2$ then

$$w(t, y(t)) = z(t) = \frac{Crt}{r - y_0}, \quad y(t) = y_0 - \frac{C\gamma_2 rt}{r - y_0}$$

which provides analyticity of the solution and concludes the proof. If $(\gamma_2 - \gamma_1) = (\ell - 2)m > 0$, then

$$\begin{aligned} Crt &= \frac{1}{2}\left((\gamma_2 - \gamma_1)z(t)^2 + 2(r - y_0)z(t)\right) \\ &= \frac{1}{2}\left((\gamma_2 - \gamma_1)z(t)^2 + 2(r - y(t))z(t) + 2\gamma_2 z(t)^2\right) \\ &= -\frac{\gamma_1 + \gamma_2}{2} z(t)^2 + (r - y(t))z(t) \end{aligned}$$

from which we deduce immediately (using $z(0) = 0$ to decide on the root)

$$z(t) = w(t, y(t)) = \frac{1}{\gamma_1 + \gamma_2}\left((r - y(t)) - \sqrt{(r - y(t))^2 - 2(\gamma_1 + \gamma_2)Crt}\right)$$

which gives

$$w(t, y) = \frac{1}{\ell m}\left((r - y) - \sqrt{(r - y)^2 - 2\ell mCrt}\right)$$

[5] The first time where this maps stops being invertible is called a *caustic* of *shock wave* depending on the context, and will be studied in details in the chapter on hyperbolic equations.

and concludes the proof.

EXERCISE 17. *Check that the previous formula for w indeed provides a solution.*

4. Examples, counter-examples, and basic classification

4.1. Failure of the Cauchy-Kovalevaskaya Theorem and evolution problems. If we consider the heat equation with initial conditions (this counter-example is due to S. Kovalevskaya)

$$\partial_t u = \partial_x^2 u, \quad u = u(t,x), \ (t,x) \in \mathbb{R}^2$$

around the point $(0,0)$, with the initial condition

$$u(0,x) = g(x)$$

we have, in the previous setting, $\Gamma = \{t = 0\}$, and the normal unit vector in $\mathbb{R}^2$ is simply $(1,0) = \mathbf{e}_1$. The non-characteristic condition writes $a_{k,0} \neq 0$, where k is the order of the equation (here $k = 2$), which is not true here. Hence the initial value problem for the heat equation is characteristic. This reflects the fact that the equation cannot be reversed in time, or in other words, the Cauchy problem is ill-posed for negative times. In particular, consider the following initial data (considered by S. Kovalevskaya)

$$g(x) = \frac{1}{1+x^2}$$

which are clearly analytic. Then let us search for an analytic solution

$$u(t,x) = \sum_{m,n \geq 0} a_{m,n} \frac{t^m}{m!} \frac{x^n}{n!}.$$

Then the PDE imposes the following relation on the coefficients

$$\forall\, m,n \geq 0, \quad a_{m+1,n} = a_{m,n+2},$$

with the initialization

$$\forall\, n \geq 0, \quad a_{0,2n+1} = 0, \quad a_{0,2n} = (-1)^n (2n)!$$

We deduce that

$$\forall\, m,n \geq 0, \quad a_{m,2n+1} = 0$$

using the inductive relation, and then

$$\forall\, m,n \geq 0, \quad a_{m,2n} = (-1)^{m+n} (2(m+n))!$$

Now since

$$\frac{(2(m+n))!}{m!(2n)!} = \frac{(4n)!}{n!(2n)!} \sim C n^{1/2} \beta^n n^n \longrightarrow +\infty$$

(using the Stirling formula) as $m = n \to \infty$, in a way which cannot be damped by geometric factors $t^n x^{2n}$, and we deduce that the entire series defining u has a radius of convergence equal to zero.

In words, what we have exploited in this proof is that the equation implies $\partial_t^k u = \partial_x^{2k} u$ for all $k \in \mathbb{N}$, and the strongest bound on the x-derivatives for general analytic initial data $u(0, x)$ are of the form $\mathrm{cst}(2k)!/r^k$, whereas on the LHS the t-derivatives should grow at most, in order to recover analyticity, as $\mathrm{cst}k!/\rho^k$, and these two things are contradictory. Hence by equating more spatial derivatives on the right hand side with less derivatives on the left hand side, one generates faster growth in the right hand side than is allowed for the left hand side to be analytic.

This example shows how the notion of characteristic boundary condition highlights some key physical and mathematical aspects of the equation at hand. It can easily be seen that a necessary and sufficient conditions for an *evolution problem*

$$\partial_t^k u = \sum_{|\alpha|=l} a_\alpha \partial_x^\alpha u$$

is that $l \le k$.

Exercise 18. *Check the last point, and formulate a similar conditions for systems.*

Hint*: For k_i time derivatives on the i-th component, no spatial derivatives on this component should be of order higher than k_i.*

4.2. Principal symbol and characteristic form. Let P be a scalar differential operator of order k:

$$Pu := \sum_{|\alpha| \le k} a_\alpha(x) \partial_x^\alpha u, \quad u = u(x), \ x \in \mathbb{R}^\ell.$$

For convenience let us assume here that the $a_\alpha(x)$ are smooth functions.

Then the *total symbol* of the operator is defined as

$$\sigma(x, \xi) := \sum_{|\alpha| \le k} a_\alpha(x) \xi^\alpha, \quad \xi^\alpha := \xi_1^{\alpha_1} \dots \xi_\ell^{\alpha_\ell}$$

and the *principal symbol* of the operator is defined as

$$\sigma_p(x, \xi) := \sum_{|\alpha|=k} a_\alpha(x) \xi^\alpha,$$

where $\xi \in \mathbb{C}$. This principal symbol is also called the *characteristic form* of the equation.

REMARK 4.1. *For geometric PDE's in an open set $\mathcal{U}$ of some manifold $\mathcal{M}$, the principal symbol is better thought of as a function on the cotangent bundle: $\sigma_p : T^*\mathcal{U} \to \mathbb{R}$.*

EXERCISE 19. *Show that the principal symbol is an* homogeneous function of degree k *in ξ, i.e.*

$$\sigma_p(x, \lambda\xi) = \lambda^k \sigma_p(x, \xi), \quad \lambda \in \mathbb{R}.$$

Then the non-characteristic condition at $x \in \Gamma$ becomes in this context

$$\sigma_p(x, \mathbf{n}(x)) \neq 0.$$

REMARK 4.2. *With a more geometric intrinsic formulation, we could say that for any $\xi \in T^*_x\mathcal{U} \setminus \{0\}$, $x \in \Gamma \subset \mathcal{U}$, with $\langle \xi, w\rangle = 0$ for all $w \in T^*_x\mathcal{U}$ tangent to Γ at $x \in \mathcal{U}$, then $\sigma_p(x, \xi) \neq 0$.*

We also introduce the *characteristic cone*[6] of the PDE at $x \in \mathbb{R}^\ell$:

$$\mathcal{C}_x := \left\{ \xi \in \mathbb{R}^\ell \ : \ \sigma_p(x, \xi) = 0 \right\}.$$

Then a surface is characteristic at a point if the normal to the surface at that point belongs to the characteristic cone at the same point.

4.3. The main linear PDE's and their characteristic surfaces. Let us go through the main linear PDE's and study their characteristic surfaces. We shall study in the next chapters some paradigmatic examples:

- Laplace's equation and Poisson's equation

$$\Delta u = 0 \quad \text{or} \quad \Delta u = f, \quad u = u(x_1, \ldots, x_\ell), \quad \left(\Delta = \sum_{j=0}^{\ell} \partial^2_{x_j} \right).$$

  In this case as we discussed the characteristic form is $\sigma_p(x, \xi) = |\xi|^2$ and the characteristic cone is $\mathcal{C}_x = \{0\}$ for any $x \in \mathbb{R}^\ell$, and any real surface cannot be characteristic to the Laplace equation. Equation without real characteristic surfaces are called *elliptic equations*.
- The wave equation

$$\Box u = 0, \quad u, \quad u = u(t, x_1, \ldots, x_\ell), \quad \left(\Box := -\partial^2_t + \sum_{j=1}^{\ell} \partial^2_{x_j} = -\partial^2_t + \Delta \right).$$

  The wave equation is obtained from the Laplace equation by the so-called *Wick rotation* $x_\ell \mapsto ix_\ell$. Its characteristic form is $\sigma_p(x, \xi) = \xi_1^2 + \cdots + \xi_{\ell-1}^2 - \xi_\ell^2$ and its characteristic cone is the

[6]The name "cone" is related to the homogeneity property described above: the characteristic cone is hence invariant by multiplication by a real number.

so-called *light cone* $\mathcal{C}_x = \{\xi_\ell^2 = \xi_1^2 + \cdots + \xi_{\ell-1}^2\}$ for any $x \in \mathbb{R}^\ell$. Any surface whose normal makes an angle $\pi/4$ with the direction $\mathbf{e}_\ell$ is a characteristic surface.

- The heat equation

$$\partial_{x_\ell} u - \Delta u = 0, \quad u = u(t, x_1, \ldots, x_\ell), \quad \left(\Delta := \sum_{j=1}^{\ell-1} \partial_{x_j}^2\right).$$

Its characteristic form is $\sigma_p(x, \xi) = \xi_1^2 + \cdots + \xi_{\ell-1}^2$ and its characteristic cone is $\mathcal{C}_x = \{\xi_1 = \cdots = \xi_{\ell-1} = 0\}$ for any $x \in \mathbb{R}^\ell$, and so the characteristic surfaces are the horizontal planes $\{x_\ell = \text{cst}\}$ (hence corresponding to an initial condition).

- The Schrödinger equation

$$i\partial_t u + \Delta u = 0, \quad u = u(t, x_1, \ldots, x_\ell) \in \mathbb{C}, \quad \left(\Delta := \sum_{j=1}^{\ell} \partial_{x_j}^2\right).$$

The Schrödinger equation is obtained from the heat equation by the *Wick rotation* $x_\ell \mapsto ix_\ell$. Its characteristic form is again $\sigma_p(x, \xi) = \xi_1^2 + \cdots + \xi_{\ell-1}^2 - \xi_\ell^2$ and its characteristic cone is again $\mathcal{C}_x = \{\xi_1 = \cdots = \xi_{\ell-1} = 0\}$ for any $x \in \mathbb{R}^\ell$, and so the characteristic surfaces are again the horizontal planes $\{x_\ell = \text{cst}\}$ (hence corresponding to an initial condition).

- The transport (including Liouville) equation

$$\sum_{j=1}^{\ell} c_j(x) \partial_{x_j} u = 0, \quad u = u(x_1, \ldots, x_\ell).$$

Then its characteristic form is

$$\sigma_p(x, \xi) = \left(\sum_{j=1}^{\ell} c_j(x) \xi_j\right)$$

and its characteristic cone is

$$\mathcal{C}_x = \mathbf{c}(x)^\perp, \quad \mathbf{c} = (c_1, \ldots, c_\ell)$$

for any $x \in \mathbb{R}^\ell$. This means that a characteristic surface is everywhere tangent to $\mathbf{c}(x)$. Then all our transport equation tells us is the behaviour of u along the characteristic surface, and what u does in the transversal direction is completely "free". This means that the existence is lost unless the initial condition on the surface satisfies certain constraints, and if a solution exists, it will not be unique. The situation is reminiscent to solving the linear system $Ax = b$ with a non-invertible square matrix A. Another viewpoint is to observe that the characteristic surfaces are the only surfaces along which two different solutions can touch each other, for if two solutions are the same on a non-characteristic surface, by uniqueness they must coincide in a neighbourhood of the surface.

Remark 4.3. *The equations all share the property that they are* linear*, and they often occur when linearising more complicated equations which play a role in Mathematical Physics, or by other types of limits.*

4.4. What is wrong with analyticity? The complex analytic setting is completely natural for the Cauchy-Kovalevskaya theorem. This is because any real analytic function uniquely extends to a complex analytic one in a neighbourhood of $\mathbb{R}^\ell$ considered as a subset of $\mathbb{C}^\ell$, and more importantly this point of view offers a better insight on the behaviour of analytic functions. Hence the complex analytic treatment contains the real analytic case as a special case. However, it is known that if we allowed only analytic solutions, we would be missing out on most of the interesting properties of partial differential equations. For instance, since analytic functions are completely determined by its values on any open set however small, it would be extremely cumbersome, if not impossible, to describe phenomena like wave propagation, in which initial data on a region of the initial surface are supposed to in uence only a specific part of spacetime. A much more natural setting for a differential equation would be to require its solutions to have just enough regularity for the equation to make sense. For example, the Laplace equation $\Delta u = 0$ already makes sense for twice differentiable functions. Actually, the solutions to the Laplace equation, i.e. harmonic functions, are automatically analytic, which has a deep mathematical reason that could not be revealed if we restricted ourselves to analytic solutions from the beginning. In fact, the solutions to the Cauchy–Riemann equations, i.e. holomorphic functions, are analytic by the same underlying reason, and complex analytic functions are nothing but functions satisfying the Cauchy–Riemann equations. From this point of view, looking for analytic solutions to a PDE in $\mathbb{R}^\ell$ would mean coupling the PDE with the Cauchy–Riemann equations and solving them simultaneously in $\mathbb{R}^{2\ell}$. In other words, if we are not assuming analyticity, $\mathbb{C}^\ell$ is better thought of as $\mathbb{R}^{2\ell}$ with an additional algebraic structure. Hence the real case is more general than the complex one, and from now on, we will be working explicitly in real spaces such as $\mathbb{R}^\ell$, unless indicated otherwise.

As soon as we allow non-analytic data and/or solutions, many interesting questions arise surrounding the Cauchy-Kovalevskaya theorem. First, assuming a setting to which the Cauchy-Kovalevskaya theorem can be applied, we can ask if there exists any (necessarily non-analytic) solution other than the solution given by the Cauchy-Kovalevskaya theorem. In other words, is the uniqueness part of the Cauchy-Kovalevskaya theorem still valid if we now allow non-analytic solutions? For linear equations an affirmative answer is given by Holmgren's uniqueness theorem. Moreover, uniqueness holds for first order equations, but fails in general for higher order equations and systems. Such a uniqueness result can also be thought of as a regularity theorem, in the sense that if u is a solution then it would be automatically analytic by uniqueness.

The second question is whether existence holds for non-analytic data, and again the answer is negative in general. A large class of counter-examples can be constructed, by using the fact that some equations, such as the Laplace and the Cauchy-Riemann equations, have only analytic solutions, therefore their initial data, as restrictions of the solutions to an analytic hypersurface, cannot be non-analytic. Hence such equations with non-analytic initial data do not have solutions. In some cases, this can be interpreted as one having "too many" initial conditions that make the problem overdetermined, since in those cases the situation can be remedied by removing some of the initial conditions. For example, with sufficiently regular closed surfaces as initial surfaces, one can remove either one of the two Cauchy data in the Laplace equation, arriving at the Dirichlet or Neumann problem. Starting with Hans Lewy's celebrated counter-example of 1957, more complicated constructions along similar lines have been made that ensure the inhomogeneous part of a linear equation to be analytic, thus exhibiting examples of linear equations with no solutions when the inhomogeneous part is non-analytic, regardless of initial data. The lesson to be learned from these examples is that the existence theory in a non-analytic setting is much more complicated than the corresponding analytic theory, and in particular one has to carefully decide on what would constitute the initial data for the particular equation.

Indeed, there is an illuminating way to detect the poor behaviour of some equations discussed in the previous paragraph with regard to the Cauchy problem, entirely from within the analytic setting, that runs as follows. Suppose that in the analytic setting, for a generic initial datum ψ it is associated the solution $u = S(\psi)$ of the equation under consideration, where $S : \psi \mapsto u$ is the solution map. Now suppose that the datum is non-analytic, say, only continuous. Then by the Weierstrass approximation theorem, for any $\varepsilon > 0$ there is a polynomial ψ_ε that is within an ε distance from ψ. Taking some sequence $\varepsilon \to 0$, if the solutions $u_\varepsilon = S(\psi_\varepsilon)$ converge locally uniformly to a function u, we could reasonably argue that u is a solution (in a generalized sense) of our equation with the (non-analytic) datum ψ. The counter-examples from the preceding paragraph suggest that in those cases the sequence u_ε cannot converge. Actually, the situation is much worse, as the following example due to J. Hadamard shows.

Consider the Cauchy problem for the Laplace equation

$$\partial_{tt}^2 u + \partial_{xx}^2 u = 0, \quad u(x,0) = a_\omega \sin(\omega x), \quad \partial_t u(x,0) = b_\omega \sin(\omega x)$$

for some parameter $\omega > 0$, whose solution is explicitely given by

$$u(x,t) = \left(a_\omega \cosh(\omega t) + \frac{b_\omega}{\omega} \right) \sin(\omega x).$$

Then if we choose $a_\omega = 1/\omega$, $b_\omega = 1$ and $\omega >> 1$, we see that the initial data is small: $u(x,0) = \mathcal{O}(1/\omega)$, $\partial_t u(x,0) = 0$, whereas the solution grows arbitrarily fast as ω tends to infinity: $u(x,1) = \sin(\omega x)(\cosh \omega)/\omega$. Hence the relation between the solution and the Cauchy data becomes more and more difficult to invert as we go to higher and higher frequencies $\omega \to \infty$. For instance if the initial data is the error of an approximation of non-analytical data in the uniform norm as $\omega \to \infty$, then the solutions with initial data given by the approximations diverge unless a_ω and b_ω decay faster than exponential. But functions that can be approximated by analytic functions with such small error form a severely restricted class, being between the smooth functions C^∞ and the analytic functions.

EXERCISE 20. *In the exercise we give a slightly amplified version of the example of J. Hadamard: consider the problem*

$$\partial_{tt}^2 u + \partial_{xx}^2 u = 0, \quad u(x,0) = \phi(x), \quad \partial_t u(x,0) = \psi(x).$$

For a given $\varepsilon > 0$ and an integer $k > 0$, construct initial data ϕ and ψ so that

$$\|\phi\|_\infty + \left\|\phi^{(1)}\right\|_\infty + \cdots + \left\|\phi^{(k)}\right\|_\infty + \|\psi\|_\infty + \left\|\psi^{(1)}\right\|_\infty + \cdots + \left\|\psi^{(k)}\right\|_\infty < \varepsilon$$

and

$$\|u(\cdot, \varepsilon)\| \geq \frac{1}{\varepsilon}.$$

Repeat the exercise with the condition on the initial data replaced by

$$\forall\, k \geq 0, \quad \left\|\phi^{(k)}\right\|_\infty + \left\|\psi^{(k)}\right\|_\infty < \varepsilon.$$

Let us contrast the previous (elliptic) example with the following (hyperbolic) one: consider the Cauchy problem for the wave equation

$$\partial_{tt}^2 u - \partial_{xx}^2 u = 0, \quad u(x,0) = \phi(x), \quad \partial_t u(x,0) = \psi(x),$$

whose solution is given by d'Alembert's formula

$$u(t,x) = \frac{\phi(x-t)+\phi(x+t)}{2} + \frac{1}{2}\int_{x-t}^{x+t} \psi(y)\,\mathrm{d}y.$$

We deduce that

$$|u(x,t)| \le \sup_{[x-t,x+t]} |\phi| + |t| \sup_{[x-t,x+t]} |\psi|$$

showing that small initial data lead to small solutions (and also showing domain of dependency). The explicit solution constructed cannot also be shown to be the unique solution by energy methods (see later in the next chapters).

This is in response to these considerations that Hadamard introduced the concept of *well-posedness* of a problem that we have introduced in the first chapter.

4.5. Basic classification. Roughly speaking, the (1) *hyperbolic*, (2) *elliptic*, (3) *parabolic*, and (4) *dispersive* classes arise as one tries to identify the equations that are similar to, and therefore can be treated by extensions of techniques developed for, the (1) wave (and transport), the (2) Laplace (and the Cauchy-Riemann), the (3) heat, and the (4) Schrödinger equations, respectively.

Indeed, the idea of *hyperbolicity* is an attempt to identify the class of PDE's for which the Cauchy-Kovalevskaya theorem can be rescued in some sense when we relax the analyticity assumption. The simplest examples of hyperbolic equations are the wave and transport equations. In contrast, trying to capture the essence of the poor behaviour of the Laplace and Cauchy-Riemann equations in relation to their Cauchy problems leads to the concept of ellipticity. Hallmarks of elliptic equations are having no real characteristic surfaces, smooth solutions for smooth data, overdeterminacy of the Cauchy data hence boundary value problems, and being associated to stationary phenomena.

The class of *parabolic equations* is a class for which the evolution problem is well-posed for positive times, but failes for negative times. The initial condition is characteristic and the Cauchy-Kovalevskaya Theorem fails. The informations is transmitted at infinite speed, and there is instanteneous regularisation: the solution becomes analytic for positive times. The latter phenomenon is extremely important and obviously cannot be captured in analytic setting.

The class of *dispersive equations* is a class which is close to transport-wave equations in the sense that their "extension" in the space-frequency phase space has the structure of a transport equation. Moreover the evolution is reversible and well-posed at the level of the linear equation, but the Cauchy-Kovalevskaya is not adapted again. And they both transport information at finite speed. However let us discuss the crucial difference between these two classes which justifies the name "dispersive".

Consider a *plane wave* solution $u(t,x) = \cos(k(t+x))$ to the wave equation

$$\partial^2_{tt} u = \partial^2_{xx} u, \quad t, x \in R,$$

with the initial data $u(0,x) = \cos(kx)$, $\partial_t u(0,x) = 0$. Then the information travels at speed 1, whatever the frequency $k \in \mathbb{R}$ of the wave. Next consider again a plane wave solution $u(t,x) = e^{i(kx-|k|^2 t)}$ to the Schrödinger equation

$$i\partial_t u + \partial^2_{xx} u = 0, \quad t, x \in R,$$

with the initial data $u(0,x) = e^{ikx}$. The information then travels at speed $|k|$ which now *depends* on the frequency! In physics words, the *dispersion relation* is $\omega(k) = \pm|k|$ for the wave equation, and $\omega(k) = -|k|^2$. In the first case, the dispersion relation is linear and there is no wave packet dispersion, while there is in the second case. This dispersive feature results in numerous mathematical consequences which are key to many Cauchy theorems. . .